

(19) 日本国特許庁 (J P)

(12) 特 許 公 報 (B 2)

(11) 特許番号

特許第3219892号
(P3219892)

(45) 発行日 平成13年10月15日 (2001. 10. 15)

(24) 登録日 平成13年 8 月10日 (2001. 8. 10)

(51) Int.Cl.⁷

識別記号

F I

G 1 0 L 21/04
13/08

G 1 0 L 3/02
3/00

A
H

請求項の数 2 (全 9 頁)

(21) 出願番号 特願平5-78098

(22) 出願日 平成 5 年 4 月 5 日 (1993. 4. 5)

(65) 公開番号 特開平6-289895

(43) 公開日 平成 6 年10月18日 (1994. 10. 18)

審査請求日 平成11年 4 月 2 日 (1999. 4. 2)

特許法第30条第1項適用申請有り 平成 4 年10月 5 ~ 7
日に高知大学にて開催された社団法人日本音響学会平成
4 年度秋季研究発表会において発表

特許権者において、実施許諾の用意がある。

(73) 特許権者 000004352

日本放送協会

東京都渋谷区神南 2 丁目 2 番 1 号

(72) 発明者 池沢 龍

東京都世田谷区砧一丁目10番11号 日本
放送協会放送技術研究所内

(72) 発明者 中村 章

東京都世田谷区砧一丁目10番11号 日本
放送協会放送技術研究所内

(72) 発明者 宮坂 栄一

東京都世田谷区砧一丁目10番11号 日本
放送協会放送技術研究所内

(74) 代理人 100083806

弁理士 三好 秀和 (外 8 名)

審査官 渡邊 聡

最終頁に続く

(54) 【発明の名称】 リアルタイム話速変換装置

1

(57) 【特許請求の範囲】

【請求項 1】 受聴音声の発声する速さ（話速）を遅く
する際、所定時間以上の無音区間を検出し、この無音区
間を発声音の息つき区間を意味するポーズ区間と判定す
るポーズ区間判定手段と、

ポーズ区間と次のポーズ区間との間をフレーズ区間と
し、このフレーズ区間の開始点から所定数の有声区間
における最高ビッチ周波数を検出する最高ビッチ周波数検
出手段と、

前記フレーズ区間の前記開始点から所定の伸張倍率、か
つ所定の減少回数に基づき、一定時間にわたって話速を
変換するとともに、一定時間の経過後においては、前記
最高ビッチ周波数を考慮した話速変換を実行する話速変
換手段と、

を備えたことを特徴とするリアルタイム話速変換装置。

2

【請求項 2】 請求項 1 に記載のリアルタイム話速変換
装置において、

前記話速変換手段は、前記一定時間の経過後における話
速を決定するに際し、前記フレーズ区間における処理対
象の有声区間の平均ビッチ周波数を求め、この平均ビッ
チ周波数と前記最高ビッチ周波数に所定の閾値を掛けた
数値との大小関係によって、処理対象の有声区間の伸張
倍率を決定することを特徴とするリアルタイム話速変換
装置。

10 【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、原音声を取り込んで、
聴覚障害者や高齢者等の音声聴取に好適なゆっくりした
速度の音声に変換するリアルタイム話速変換装置に関す
る。

【0002】[発明の概要]

本発明は、原音声を取り込んで、聴覚障害者や高齢者等の音声聴取に好適なゆっくりした速度の音声に変換するリアルタイム話速変換装置に関するものであり、リアルタイム処理で、音声のイントネーション（ピッチ周波数）の変化を検出するとともに、この検出結果に基づいてイントネーションの高い部分では、話速を緩め、低い部分では、話速を早めるという規則で話速を変化させることにより、原音声の発話時間を保ったまま、原声を聞き易い良好な音声に変換するものである。

【0003】

【従来の技術】一般に、受聴者が加齢ないしなんらかの障害などによって音声識別臨界速度（音声を正確に識別できる最大の話速）などの受聴能力が低下すると、通常の速さの音声や早口で話される音声の識別度が大幅に低下するようになる。

【0004】そして、従来においては、このような聴力障害を持つ人のための補聴手段として、補聴器しかなかった。

【0005】

【発明が解決しようとする課題】ところで、上述した補聴器は単に周波数特性の改善や利得制御などによって聴覚系の外耳、中耳の伝達特性のみを補償するための機器であることから、聴覚中枢の劣化に關与する音声の識別能力の低下を補償することができないという問題があった。

【0006】そこで、このような問題を解決する方法として、原音声の品質を保ったまま、話速を変換する技術が開発されている。

【0007】この話速変換技術では、音声の話速のみを一樣に遅くすることにより、特に高齢者や聴覚障害者等にとっては、はるかに聞き易い音声にすることが可能であるが、この操作によって音声の発話時間も必然的に伸張する。しかし、放送等では、伸張前の音声の発話時間は、決められた時間内に収まるように発話されていることから、このような音声の伸張を行なうと、上記制限時間内に収まらなくなる可能性が生じる。また、テレビジョン等のように音声と映像を同期して提供するような場合に、音声のみを伸張すると、映像との間に時間的な「ずれ」が生じ、これが聞き取りに悪影響を及ぼす虞が発生する。

【0008】このため、このような時間的な「ずれ」を考慮した話速変換技術をオフライン処理で実現するものが開発されているものの、時間的な「ずれ」を考慮した話速変換技術をリアルタイム処理で実現することができるものは未だ開発されていない。

【0009】本発明は上記の事情に鑑み、上述した時間的な「ずれ」に伴う問題点を解決するため、リアルタイム処理で、発話音声の意味上重要と考えられる部分の話速を適度に遅くし、それ以外の部分を逆に早めること

によって、発話時間を実質的に伸張させることなく、全体としてゆっくりとした聞き易い音声に変換することができるリアルタイム話速変換装置を提供することを目的としている。

【0010】

【課題を解決するための手段】上記の目的を達成するために本発明は、請求項1では、受聴音声の発声する速さ（話速）を遅くする際、所定時間以上の無音区間を検出し、この無音区間を発声音の息つき区間を意味するポーズ区間と判定するポーズ区間判定手段と、ポーズ区間と次のポーズ区間との間をフレーズ区間とし、このフレーズ区間の開始点から所定数の有音区間における最高ピッチ周波数を検出する最高ピッチ周波数検出手段と、前記フレーズ区間の前記開始点から所定の伸張倍率、かつ所定の減少関数に基づき、一定時間にわたって話速を変換するとともに、一定時間の経過後においては、前記最高ピッチ周波数を考慮した話速変換を実行する話速変換手段とを備えたことを特徴としている。請求項2では、請求項1に記載のリアルタイム話速変換装置において、前記話速変換手段は、前記一定時間の経過後における話速を決定するに際し、前記フレーズ区間における処理対象の有音区間の平均ピッチ周波数を求め、この平均ピッチ周波数と前記最高ピッチ周波数に所定の閾値を掛けた数値との大小関係によって、処理対象の有音区間の伸張倍率を決定することを特徴としている。

【0011】

【作用】上記の構成によれば、ポーズ区間判定手段は、受聴音声の発声する速さ（話速）を遅くする際、所定時間以上の無音区間を検出し、この無音区間を発声音の息つき区間を意味するポーズ区間と判定する。最高ピッチ周波数検出手段は、ポーズ区間と次のポーズ区間との間をフレーズ区間とし、このフレーズ区間の開始点から所定数の有音区間における最高ピッチ周波数を検出する。そして、話速変換手段は、前記フレーズ区間の前記開始点から所定の伸張倍率、かつ所定の減少関数に基づき、一定時間にわたって話速を変換するとともに、一定時間経過後においては、前記最高ピッチ周波数を考慮した話速変換を実行する。これにより、時間的な「ずれ」に伴う問題点を解決しながら、発話音声の意味上重要な部分の話速を適度に遅くし、それ以外の部分を逆に速めることによって、発話時間を実質的に伸張させることなく、リアルタイムで全体としてゆっくりとした聞きやすい音声に変換する。

【0012】

【実施例】《実施例の構成》

図1は本発明によるリアルタイム話速変換装置の一実施例を示すブロック図である。

【0013】この図に示すリアルタイム話速変換装置は、音声入力回路1と、CPU回路2と、PROM回路3と、入力バッファ回路4と、処理バッファ回路5と、

ファイル回路6と、音声出力回路7と、バス8とを備えており、音声入力回路1によって話速変換対象となる音声（原音声）を取り込み、リアルタイム処理で、原音声のイントネーション（ピッチ周波数）の変化を検出するとともに、この検出結果に基づいてイントネーションの高い部分では、話速を緩め、低い部分では、話速を早めるという規則で話速を変化させることにより、原音声の発話時間を保ったまま、原音声を聞き易い良好な音声に変換する。

【0014】音声入力回路1は、原音声を入力するための一般的な構成の回路、例えばマイクロホン、音調回路、アナログディジタル変換器、音声記憶再生（録音）回路、音声記憶媒体（例えば、ICメモリ、ハードディスク、フロッピーディスクまたはVTR）、およびインタフェース回路等を備えており、話速変換対象となる音声を取り込み、これをディジタル形式の音声信号に変換するとともに、CPU回路2からの指示に基づいてフレーム単位で入力バッファ回路4に供給する。

【0015】入力バッファ回路4は、必要な容量のRAMなどによって構成され、CPU回路2の作業域として使用される部分であり、音声入力回路1から出力される音声信号を取り込んでこれを記憶するとともに、CPU回路2からの指示に基づいて記憶している音声信号を処理バッファ回路5に転送する。

【0016】処理バッファ回路5は、必要な容量のRAMなどによって構成され、CPU回路2の作業域として使用される部分であり、入力バッファ回路4から出力される音声信号を取り込んでこれを記憶するとともに、CPU回路2からの指示に基づいて記憶している音声信号をファイル回路6などに転送する。

【0017】ファイル回路6は、RAMの他に、ICメモリやフロッピーディスク等の音声記憶媒体によって構成され、本発明に係わる有声区間の伸張された音声信号と、無音区間の短縮の処理を施された信号などを格納するメモリであり、処理バッファ回路5から処理済みの音声信号が出力されたとき、これを取り込んで記憶し、その後CPU回路2からの指示に基づいて記憶している音声信号を音声出力回路7に供給する。

【0018】音声出力回路7は、ファイル回路6内の音声信号を外部に出力するための一般的な構成の回路、例えばインタフェース回路、ディジタルアナログ変換器、スピーカー、録音装置（あるいは放送機器）等を備えており、ファイル回路6から音声信号が出力されたとき、これを取り込んで音声に変換しながら、外部に出力する。

【0019】また、CPU回路2は、ワンチップマイクロコンピュータ等によって構成される部分であり、PROM回路3に格納されているプログラムに基づいて装置全体の制御や各種のデータ処理を行なう。

【0020】また、PROM回路3は、CPU回路2の

動作を規定するプログラムや各種の処理で使用される定数データなどの格納場所として使用される部分であり、CPU回路2からの読出し指令に応じて記憶しているプログラムや定数データを読出してCPU回路2に供給する。

【0021】《実施例の動作》次に、図1に示すブロック図および図2、図3に示すフローチャート、図4に示すタイミング図を参照しながら、この実施例の動作を説明する。

【0022】まず、CPU回路2は最初に音声入力回路1に入力されて処理された音声信号をフレームと呼ばれる一定長、例えば3.3ms毎に切出し、これを入力バッファ回路4に転送させて格納させる（ステップST1）。

【0023】この後、CPU回路2は入力バッファ回路4に格納されている音声信号を各フレーム毎に自己相関法や零クロス法などの方法で処理して各フレーム毎に有声、無声、無音の判定を行なう。但し、この場合、人が発する有声および無声以外の入力音（例えば、低レベルの雑音や背景音など）については、原則として無音として処理する（ステップST2）。

【0024】次いで、CPU回路2は今回のフレームについての有声、無声、無音の判定結果と、前回のフレームについての有声、無声、無音の判定結果とが同じであるかどうかを判定し（ステップST3）、これらが同じ種類であれば、上述したフレームの切出し処理に戻って同じ処理を繰り返し、また違う種類、例えば前回のフレームが有声区間であり、今回のフレームが無声区間であれば、それまで同じ種類の区間と判定されている音声信号を処理バッファ回路5に転送して格納させる（ステップST4）。

【0025】これによって、図4に示す如く、音声入力回路1によって取り込まれた音声が有声区間と、無声区間と、無音区間とに区分されて処理バッファ回路5に格納される。

【0026】この後、CPU回路2は処理バッファ回路5に格納されている各音声信号のうち、無音区間と判定された区間の中で、その区間長が250ms以上の無音区間をポーズ区間（発声音の息つき区間）と判定するとともに、各ポーズ区間の間にある区間をフレーズ区間（一息で発声される区間）と判定する（ステップST5）。

【0027】そして、CPU回路2は各フレーズ区間の有声区間と判定された区間に対し、図3に示す有声区間処理10を行なう（ステップST6）。

【0028】この有声区間処理では、CPU回路2は最初に、処理対象となっている有声区間がポーズ区間直後の有声区間かどうかを判定し（ステップST15）、ポーズ区間直後の有声区間であれば、フレーズ区間の開始点（Phst）から3つの有声区間（第1有声区間、

10

20

30

40

50

第2有声区間、第3有声区間)を抽出してこれら第1有声区間～第3有声区間の各ピッチ周波数のうち、最高のピッチ周波数を最高ピッチ周波数Pitch_maxとするとともに、第1有声区間の開始点V_stにおける話速の伸張倍率を“rs”とする(ステップST16)。

【0029】この後、CPU回路2は処理対象となる音声信号が第1有声区間の開始点V_stから予め設定さ*

$$f(t) = re + (1/2) \cdot (rs - re) \cdot \{\cos \pi \cdot (t - V_st) / T + 1.0\} \quad \dots (1)$$

但し、 $t: t = V_st \sim V_st + T$

また、このとき、この範囲では、無音区間および無声区間に対し、何等の処理も施さない。

【0031】また、処理対象となる音声信号が第1有声区間の開始点V_stから予め設定されている長さの時間Tを経過していれば(ステップST17)、CPU回*

$$Pitch(n) > Pitch_max \times Th2 \quad \dots (2)$$

但し、Th2:しきい値であり、この実施例では、 $Th2 = 0.7$ 。

【0033】そして、第n音声区間が上記(2)式を満たしていれば、CPU回路2はこの第n音声区間の開始点を“V2_st”とし(ステップST20)、伸張倍率を“rs-Th3”と設定する。上述した期間内かどうかの判定処理および減少関数f(t)を使用した有声区間の伸張処理を行ない、開始点V2_stから期間Tまでの範囲で、話速の伸張倍率を“rs-Th3”から“re”まで変化させる(ステップST18)。

【0034】この場合、この実施例では、しきい値Th3の値を“0.1”に設定している。

【0035】また、前記(2)式が満たされていないければ(ステップST19)、CPU回路2は有声区間の伸張倍率re、すなわち話速を最も速い状態のままにする(ステップST21)。

【0036】以下、CPU回路2は次のポーズ区間まで、有声区間が検出される毎に、この有声区間内の音声信号に対して上述した処理を繰り返し行なう。

【0037】そして、この処理が終了した後、CPU回路2は処理バッファ内にある話速変換済みの音声信号をファイル回路6に転送させて格納させるとともに、処理バッファ回路5をクリアする(ステップST7)。

【0038】また、上述した識別区間処理において(ステップST5)、処理対象となる区間が無声区間と判定されれば、CPU回路2はこの区間の音声信号を処理バッファ回路5からファイル回路6に転送させて格納させた後、処理バッファ回路5をクリアする(ステップST8)。

【0039】また、上述した識別区間処理において(ステップST5)、処理対象となる区間が無音区間と判定

＊れている長さの時間T(この実施例では、2000ms)が経過したかどうかを判定し(ステップST17)、時間Tが経過していなければ、話速の伸張倍率を予め設定されている適切な減少関数、例えば次式に示す余弦関数f(t)を用いて“rs”から“re”まで変化させる(ステップST18)。

【0030】

※路2は処理対象となっている音声信号を含む区間(第n音声区間)(但し、 $n \geq k$)における平均ピッチ周波数Pitch(n)が次式を満たすかどうかを判定する(ステップST19)。

【0032】

されれば、CPU回路2はこの区間がポーズ区間かどうか判定し(ステップST9)、ポーズ区間であるときには、文章と文章との区切れ(句点)と判断して、文章を聴感上の違和感なく最短に短縮するため、予め設定されているアルゴリズムの短縮処理を行なって無音区間を短縮する(ステップST10)。

【0040】また、上述した識別区間処理において(ステップST5)、無音区間と判定されても、ポーズ区間でなければ(ステップST9)、CPU回路2はこの短縮処理をスキップする。

【0041】この後、CPU回路2は処理バッファ回路5内にある処理済みの無音区間の信号をファイル回路6に転送させて格納させた後、処理バッファ回路5をクリアする(ステップST11)。

【0042】以下、CPU回路2は処理対象となる音声信号が無くなるまで(ステップST12)、上述した処理を繰り返し行なう。

【0043】また、上述した処理と並行して、CPU回路2はファイル回路6内に格納されている処理済みの音声信号を音声出力回路7に転送させて音声として出力させる。

【0044】《実験例》そして、表1に示す実際のニュース音声を含む音声を上述したリアルタイム話速変換装置で処理したとき、文章中の(1)-1～(1)-6に対し、各々の区切れをポーズと認識して各フレーズの開始点で話速を遅くすることができ、(1)-1、(1)-4の後半部分(下線部分)で話速を遅くすることができた。

【0045】

【表1】

- (1)-1 神奈川県の大和市と綾瀬市にまたがる厚木基地の周辺住民が
 -2 基地を使用する米軍機などの騒音によって
 -3 積極的、肉体的に大きな被害を受けたとして、
 -4 国を相手におこしていた第2次厚木基地訴訟の最終弁論が、
 -5 今日、横浜地方裁判所で行われ、
 -6 提訴以来7(七)年にわたって争われていた裁判が結審しました。

そして、この実験において、話速の伸張倍率を“rs = 10＊秒）にすることができた。

1. 25”、“re = 0. 9”としたとき、136秒の長さの原音声を136秒の長さの音声にすることができ、“rs = re = 1. 3”と話速を一律に遅くした場合と比べて、吸収率αを100％（未吸収時間0. 0 ＊

【0046】この場合、吸収率αは次式で表わされる値である。

【0047】

$$\alpha = \{ (T1 - T2) / T1 \} \cdot 100 \quad \dots (3)$$

但し、T1：話速を一律の場合の伸張時間

T2：話速を変化させた場合の伸張時間

したがって、この実施例を使用することにより、文章間の無音区間を効果的に短縮し、これによって全体の時間長を伸張せずに、リアルタイムで原音声をゆっくりした音声に変換することができる。

【0048】このようにこの実施例においては、音声入力回路1によって話速変換対象となる音声（原音声）を取り込み、リアルタイム処理で、原音声のイントネーション（ピッチ周波数）の変化を検出するとともに、この検出結果に基づいてイントネーションの高い部分では、話速を緩め、低い部分では、話速を早めるという規則で話速を変化させることにより、原音声の発話時間を保ったまま、原音声を聞き易い良好な音声に変換するようにしたので、発話音声の意味上重要な部分の話速は適度に遅くし、それ以外の部分は逆に速めることができ、これによって発話時間を実質的に伸張させることなく、全体としてゆっくりとした聞きやすい音声に変換することができる。

【0049】

【発明の効果】以上説明したように本発明によれば、時間的な「ずれ」に伴う問題点を解決するため、リアルタイム処理で、発話音声の意味上重要な部分と考えられ

る部分の話速を適度に遅くし、それ以外の部分を逆に速めることによって、発話時間を実質的に伸張させることなく、全体としてゆっくりとした聞きやすい音声に変換することができる。

20 【図面の簡単な説明】

【図1】本発明によるリアルタイム話速変換装置の一実施例を示すブロック図である。

【図2】図1に示すリアルタイム話速変換装置の動作例を示すメインフローチャートである。

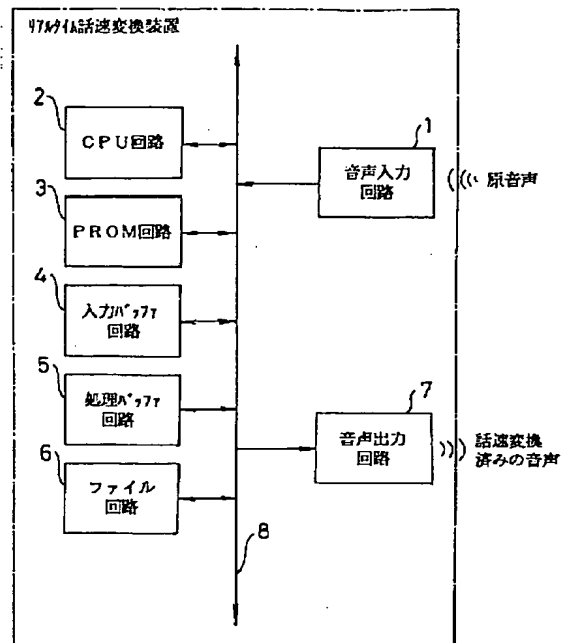
【図3】図1に示すリアルタイム話速変換装置の動作例を示す有聲区間処理ルーチンの一例を示すフローチャートである。

【図4】図1に示すリアルタイム話速変換装置の動作例を示すタイミング図である。

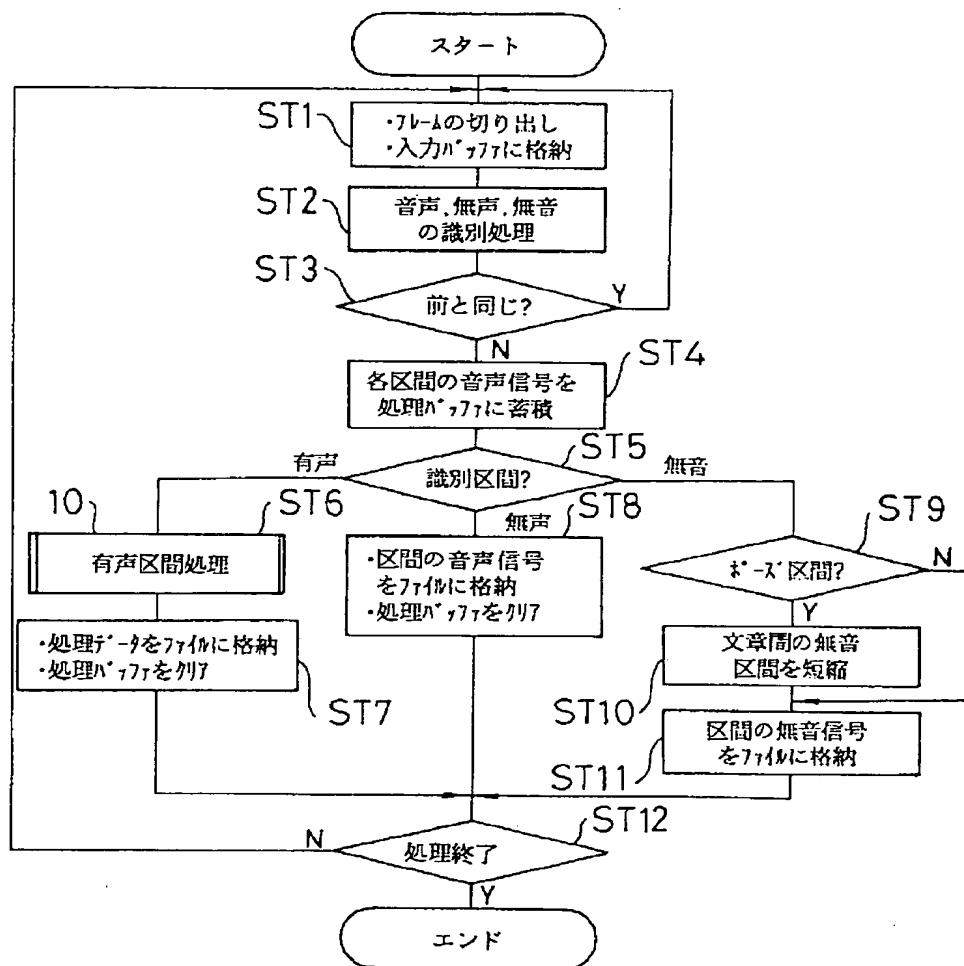
30 【符号の説明】

- 1 音声入力回路
- 2 CPU回路
- 3 PROM回路
- 4 入力バッファ回路
- 5 処理バッファ回路
- 6 ファイル回路
- 7 音声出力回路
- 8 バス

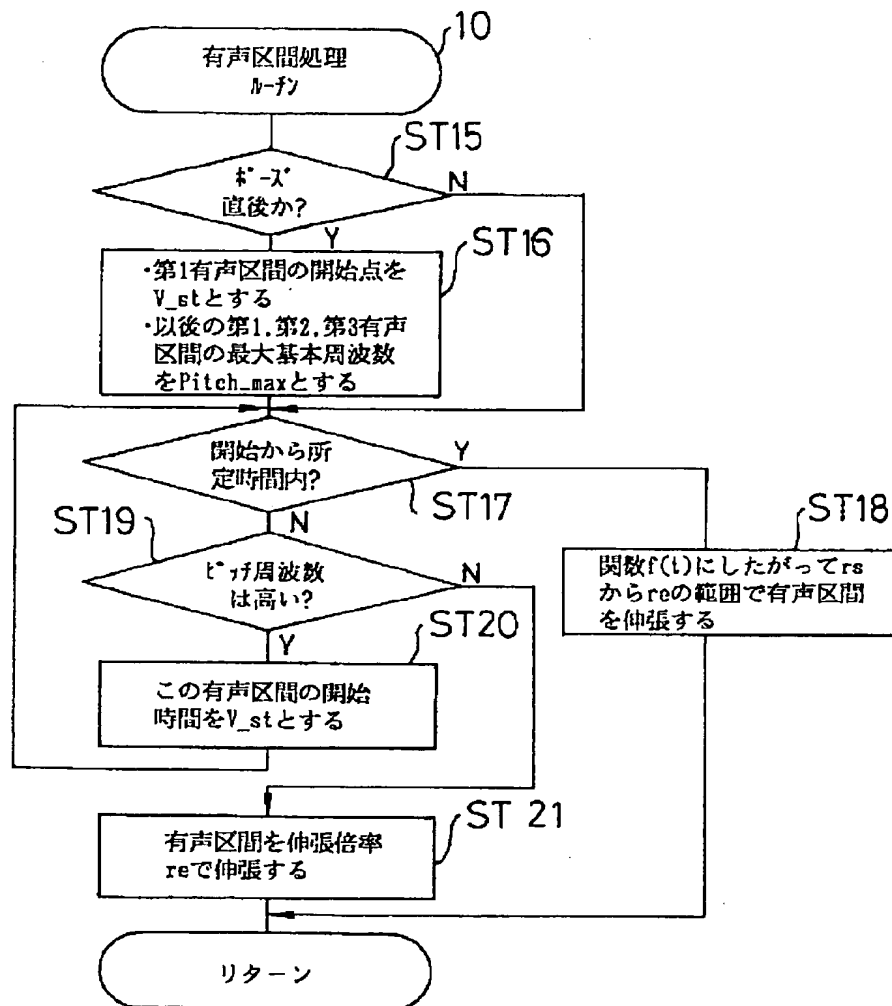
【図1】



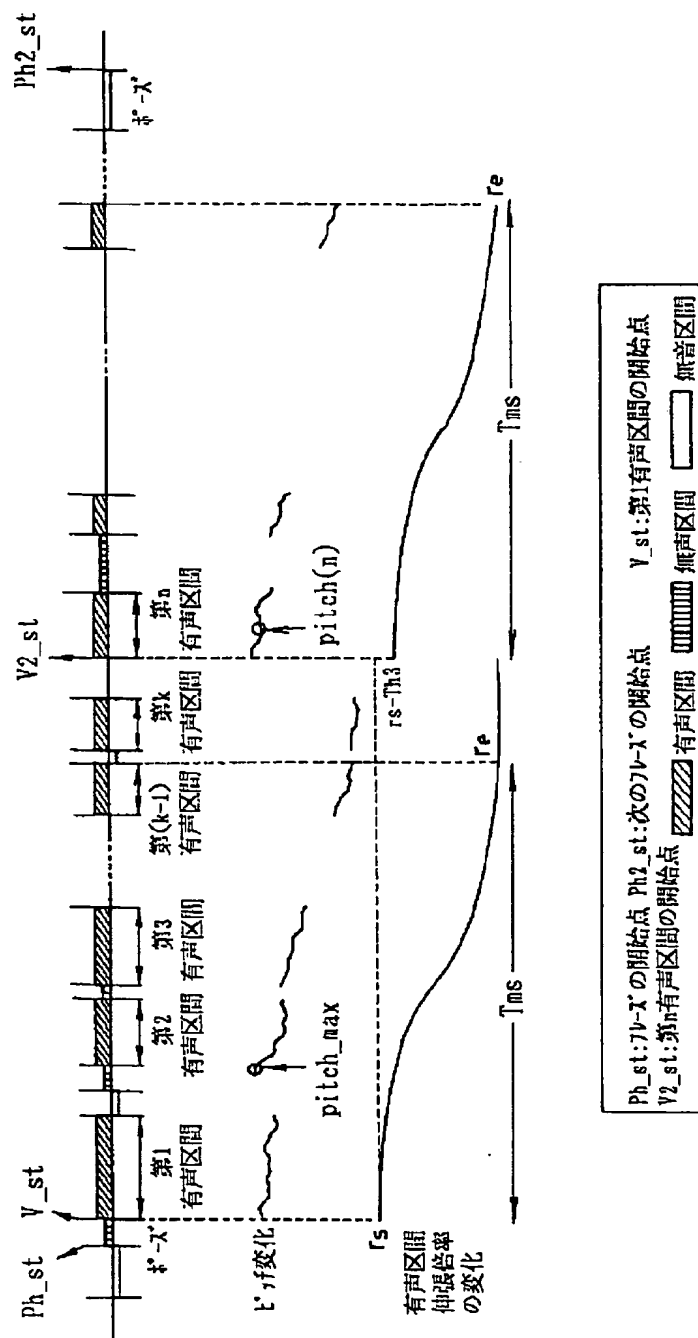
【図2】



【図3】



【図4】



フロントページの続き

(58)調査した分野(Int.Cl.⁷, D B名)

G10L 21/04